REMARKS

With this amendment, claims 14, 17, 20, 22, 39, 42, 45-47, and 58 have been amended for clarity. In addition, URLs in the specification have been amended in such a manner that they have been deactivated. No new matter has been added by virtue of these claim amendments or the amendments to the specification. Upon entry of the present amendments, claims 14, 15, 17, 20-22, 39, 40, 42, 45-47, and 58 will remain pending in the above-identified application.

THE OBJECTIONS TO THE SPECIFICATION SHOULD BE WITHDRAWN

In the July 20, 2005 Office Action, the Examiner objected to embedded hyperlinks that appear on pages 12, 13, and 23 of the specification. In response, Applicants have amended these hyperlinks in such a manner that they are inactivated. Applicants have stripped the "http://" designation from each of the links. In this way, the specification still discloses the hyperlink addresses, but does so in a manner that will not accidentally lead to active hyperlinks in the specification when the patent application issues and is published on the USPTO web site. Accordingly, Applicants request that the objection to the specification be withdrawn.

THE 35. U.S.C. § 112 SECOND PARAGRAPH REJECTION SHOULD BE WITHDRAWN

The Examiner has rejected claims 14, 15, 17, 20-22, 39, 40, 42, 45-47, and 58 under 35 U.S.C. § 112, second paragraph, for lack of antecedent basis of the phrase "the correlation value associated with the respective genotypic data structures" in independent claims 14, 17, 20, 22, 39, 42, 45-47, and 58. Claims 15, 21, and 40 are rejected for being dependent from rejected claim 14, 20, or 39. In response, Applicants have amended each of these independent claims in the same manner. As an example, claim 14 has been amended in relevant part as follows:

> repeating said establishing and determining steps for each locus in said plurality of loci, thereby establishing a plurality of genotypic data structures and, for each <u>respective</u> genotypic data structure in the plurality of genotypic data structures, ~~an associated~~ <u>determining a</u> correlation value;
>
> identifying one or more genotypic data structures in said plurality of genotypic data structures, wherein <u>the correlation value for</u> each respective genotypic data structure in the one or more genotypic data structures ~~has the~~

~~property that the correlation value associated with the respective genotypic data structure forms~~ is a high correlation value relative to the correlation values of genotypic data structures in said plurality of genotypic data structures that are not in said one or more genotypic data structures; wherein the loci that correspond to said one or more genotypic data structures represent said one or more candidate chromosomal regions that associate with said phenotype and wherein an amount of said genome that is included in each locus in said plurality of loci is predetermined;

Each of the independent claims has been amended in this manner.

With the above-described claim amendments, it is now clear that the repeating step of each of the independent claims causes a correlation value to be determined for each genotypic data structure in a plurality of data structures. The claimed repeating step is supported by Fig. 2 of the specification, in particular steps 206 and 208 coupled with the step 212 of Fig. 2. In steps 206 and 208 a genotypic data structure is established and a correlation value is determined for the genotypic structure data structure as described in the specification, for example, on page 23, line 16, through page 26, line 16. Step 212, the repeating step in which the establishing and determining steps are repeated for a different genotypic data structure, is described, for example, in the specification on page 26, line 21, through page 27, line 15. Thus, in each of the independent claims, as amended, a correlation value is determined for each genotypic data structure in a plurality of genotypic data structures by operation of the repetition of the establishing (Fig. 2, step 206) and determining steps (Fig. 2, step 208) in the same manner illustrated in Fig. 2 of the specification.

The identifying step of each of the independent claims is supported by steps 214 and 216 of Fig. 2 of the specification. Steps 214 and 216 are described, for example, on page 27, line 26, through page 27, line 17, of the specification. Here, consistent with the independent claims as amended, the specification teaches that genotypic data structures that have high correlation values are selected.

The Examiner has further rejected claims 14, 15, 17, 20-22, 39, 40, 42, 45-47, and 58 under 35 U.S.C. § 112, second paragraph, for lack of antecedent basis for reciting the phrase "the property" in independent claims 14, 17, 20, 22, 39, 42, 45-47, and 58. Claims 15, 21, and 40 are rejected for being dependent from rejected claim 14, 20, or 39. In response, Applicants have amended these independent claims to cancel the phrase "the property" and to amend the phrase "the associated correlation value" in favor "determining a correlation value."

In view of the above-identified claim amendments, Applicants believe that each of the 35 U.S.C. § 112, second paragraph, rejections raised by the Examiner have now been obviated. Accordingly, Applicants respectfully request that the 35 U.S.C. § 112, second paragraph, rejection of the pending claims be withdrawn.

THE 35. U.S.C. § 112 FIRST PARAGRAPH REJECTION SHOULD BE WITHDRAWN

The Examiner has rejected claims 14, 15, 17, 20-22, 39, 40, 42, 45-47, and 58 under 35 U.S.C. § 112, first paragraph, for allegedly failing to comply with the written description requirement. In particular, the Examiner contends that the claim limitation "the one or more genotypic data structures has the property that the correlation value associated with the respective genotypic data structure" is not found in the instant specification. Applicants have amended this claim limitation in the manner described above in the section addressing the 35 U.S.C. § 112, second paragraph, rejections. Specifically, the following claim amendment has been made:

> wherein the correlation value for each respective genotypic data structure in the one or more genotypic data structures ~~has the property that the correlation value associated with the respective genotypic data structure forms~~ is a high correlation value relative to the correlation values of genotypic data structures in said plurality of genotypic data structures that are not in said one or more genotypic data structures

This claim limitation, as amended, is supported by steps 214 and 216 of Fig. 2 of the instant specification. Steps 214 and 216 are described, for example, on page 27, line 26, through page 27, line 17, of the specification. Here, consistent with the independent claims as amended, the specification teaches that genotypic data structures that have high correlation values are selected. Additional support for this claim limitation is found on page 5, lines 1-10 of the specification reproduced for the Examiner's convenience:

> The phenotypic and genotypic data structures are then compared to form a correlation value. The process continues with the establishment of another genotypic data structure that corresponds to a different loci and the concomitant comparison of this genotypic data structure to the phenotypic

> structure until several of the loci in the genome of the organism have been tested in this manner. In this way, one or more genotypic data structures are identified that form a high correlation value relative to all other genotypic data structures that have been compared to the phenotypic data structure. Further, the loci in the genome of the organism that correspond to the highly correlated genotypic data structures represent one or more candidate chromosomal regions that may be associated with the phenotype of interest.

This claim passage illustrates how one or more genotypic data structures are identified that each form a high correlation value relative to all other genotypic data structures, consistent with the limitation identified by the Examiner. Thus, Applicants contend that the claim limitation identified by the Examiner is fully supported by the specification in its amended form.

The Examiner has rejected claims 14, 15, 17, 20-22, 39, 40, 42, 45-47, and 58 under 35 U.S.C. § 112, first paragraph, on the additional basis that the claim limitation "plurality of genotype data structures that are not in said one or more genotypic data structures" is not found in the instant specification. Applicants have not amended this claim limitation in the instant response because Applicants respectfully disagree with the Examiner on this point. A plurality of genotypic data structures are established in the instant application and in each of the independent claims. This plurality of genotypic data structures is then divided into two classes: (i) the identified one or more genotypic data structures that have high correlation values and (ii) those genotypic data structures in the plurality of genotypic data structures that are not in the identified one or more genotypic data structures (*i.e.,* that do not have high correlation values).

Applicants have noted several passages in the specification where it is clear that one or more genotypic data structures are selected from a plurality of genotypic data structures. For example, on page 5, lines 1-10, of the specification, reproduced above, "one or more genotypic data structures are identified that form a high correlation value relative to all other genotypic data structures that have been compared to the phenotypic data structure." This claim limitation is further supported by step 216 of Fig. 2. Page 28, lines 8-9, of the specification states that, in processing step 216, the genotypic data structures that achieve the highest correlation values are selected. The specification makes it very clear that these genotypic data structures that achieve the highest correlation values are relative to those genotypic data structures not selected. For example, on page 28, lines 12-14, the specification states that "[i]n one embodiment, the selection process in processing step 216 is

performed by selecting genotypic data structures that form a correlation value that is a predetermined number of standard deviations above the mean correlation value.

In view of the above-identified claim amendments, Applicants believe that each of the 35 U.S.C. § 112, first paragraph, rejections raised by the Examiner have now been obviated. Accordingly, Applicants respectfully request that the 35 U.S.C. § 112, first paragraph, rejection of the pending claims be withdrawn.

THE 35. U.S.C. § 101 REJECTION SHOULD BE WITHDRAWN

The Examiner has rejected claims 14, 15, 17, 20-22, 39, 40, 42, 45-47, and 58 under 35 U.S.C. § 101 because the claimed invention is allegedly directed to non-statutory algorithm type subject matter. For the reasons discussed below, Applicants respectfully traverse the rejection.

The rejection of the claims under 35 U.S.C. § 101 in the July 20, 2005 Office Action represents the second time the claims have been rejected on the very same basis. The first time the claims were rejected on this basis was in the October 19, 2004 Office Action, where, on page 3, the Examiner stated that the claims were rejected because they are "directed to a method comprising algorithmic steps for manipulating genotypic and phenotypic data without any physical alteration step, which is considered to be non-statutory subject matter." This per se data transformation test of Applicants' claims is in complete contradiction to the instructions provided by John J. Doll, Commissioner for Patents, in the October 26, 2005, *Interim Guidelines for Examination of Patent Applications for Patent Subject Matter Eligibility* (hereinafter "*Interim Guidelines*").[1] Page 20 of the *Interim Guidelines* explains:

> b. Practical Application That Produces a Useful, Concrete, and Tangible Result
>
> For eligibility analysis, physical transformation "is not an invariable requirement, but merely one example of how a mathematical algorithm [or law of nature] may bring about a useful application." AT&T, 172 F.3d at 1358-59, 50 USPQ2d at 1452. If the examiner determines that the claim does not entail the transformation of an article, then the examiner shall review the claim to

[1] A copy of the *Interim Guidelines* is attached as Exhibit A. Specifically, page 42 of the *Interim Guidelines* states that a per se data transformation test is not to be applied by Examiners in determining whether the claimed invention is patent eligible subject matter.

> determine if the claim provides a practical application that produces a useful, tangible and concrete result.

Applicants respectfully submit that the rejected claims each provides a useful, concrete, and tangible result.

Applicants' claims provide a useful result. According to page 20 of the *Interim Guidelines*, a claim provides a useful result when it is: (i) specific, (ii) substantial, and (iii) credible. Applicants' claims are for specific methods, computer program products, and computer systems that associate a phenotype with one or more chromosomal regions (termed "candidate chromosomal regions") in a genome of a species. This utility is specifically claimed in each of the independent claims. For example, the preamble of claim 14 begins:

> A method of associating a phenotype with one or more candidate chromosomal regions

and, in the claim body, positively recites the limitation:

> wherein the loci that correspond to said one or more genotypic data structures represent said one or more candidate chromosomal regions that associate with said phenotype.

Thus, through the claimed identification of genotypic data structures that have high correlation scores, loci that correspond to such genotypic data structures are identified as the one or more candidate chromosomal regions that associate with a phenotype. The association of a phenotype with one or more candidate chromosomal regions is a substantial utility because it helps to identify the genes that are responsible for that particular phenotype, and is credible. This is evidenced by the reception of Applicants' work in the scientific community. For example, Applicants' claimed invention is reported in Grupe *et al.*, 2001, *Science* 292:1915-1918,[2] the abstract of which states:

> Experimental murine genetic models of complex human disease show great potential for understanding human disease pathogenesis. To reduce the time required for analysis of such models from many months down to milliseconds, a computational method for predicting chromosomal regions regulating phenotypic traits and a murine database of single nucleotide polymorphisms were developed. After entry of phenotypic information obtained from inbred

[2] Grupe *et al.*, 2001, Science 292:1915-1918 is enclosed as Exhibit B.

> mouse strains, the phenotypic and genotypic information is analyzed in silico to predict the chromosomal regions regulating the phenotypic trait.

As the *Science* abstract explains, using Applicants' claimed invention, genetic models can be rapidly analyzed. Thus, Applicants' claims provide a useful result because they are directed to an invention that is (i) specific, (ii) substantial, and (iii) credible.

Applicants' claims provide a concrete result. According to page 22 of the *Interim Guidelines*, the question of whether a claim provides a "concrete" arises when a result cannot be assured. Thus, the requirement for providing a concrete result is not a physical alteration requirement. Rather, as noted on page 22 of the *Interim Guidelines*, the test for concreteness is satisfied when the result is substantially repeatable. As further noted in the *Interim Guidelines*, the opposite of "concrete" is unrepeatable or unpredictable. Applicants' claims provide highly repeatable results, as evidenced by the experimental results in Applicants' specification and the fact that representative work in accordance with Applicants' claims is published in the reputable peer-reviewed journal *Science*.

Applicants' claims provide a tangible result. According to current USPTO policy, the "tangible result" requirement is not a physical alteration requirement: "[t]he tangible requirement does not necessarily mean that a claim must either be tied to a particular machine or apparatus or must operate to change articles or materials to a different state or thing." (*Interim Guidelines*, p. 21.) All that is required is to set forth a practical application. As discussed above, Applicants have satisfied this requirement. Applicants have identified methods, computer program products, and computer systems that have the practical application of associating a phenotype with one or more chromosomal regions (termed "candidate chromosomal regions") in a genome of a species. The claimed invention can reduce the time required for analysis of genetic models from many months down to milliseconds.

Summary. As the above discussion explains, Applicants claims are directed to a practical application that produces a useful, tangible, and concrete result. Accordingly, Applicants respectfully request that the 35 U.S.C. § 101 rejection of claims 14, 15, 17, 20-22, 39, 40, 42, 45-47, and 58 be withdrawn.

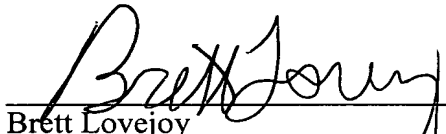
CONCLUSION

In view of the above remarks, Applicants respectfully submit that the subject application is in good and proper order for allowance. Withdrawal of the Examiner's rejections and objections and early notification to this effect are earnestly solicited.

No fee is believed owed in connection with filing of this amendment and response. However, should the Commissioner determine otherwise, the Commissioner is authorized to charge any underpayment or credit any overpayment to Jones Day Deposit Account No. 50-3013 for the appropriate amount. A copy of this sheet is attached.

Respectfully submitted,

Date: December 19, 2005



42,813
(Reg. No.)

Brett Lovejoy
JONES DAY
222 East 41st Street
New York, New York 10017-6702
(415) 875-5744

EXHIBIT A

OCTOBER 26, 2005, INTERIM GUIDELINES FOR EXAMINATION OF PATENT APPLICATIONS FOR PATENT SUBJECT MATTER ELIGIBILITY

EXHIBIT B

GRUPE *ET AL.*, 2001, SCIENCE 292:1915-1918

of MDM2 (*14*) (Fig. 2), inhibition of MDM2 binding, although preventing p53 degradation, would not block p53 nuclear export and thus would not efficiently accumulate p53 in the nucleus to allow maximal p53 activation. On the other hand, inhibiting p53 nuclear export without breaking its binding with MDM2, although causing the nuclear accumulation of p53, would not reach maximal p53 activation either because MDM2, in addition to its activity in promoting cytoplasmic p53 degradation, can also directly inhibit p53's transactivating activity in the nucleus (*4*). We suggest that DNA damage–induced phosphorylation may achieve optimal p53 activation through the additive and complementary action of both inhibiting MDM2 binding to, and the nuclear export of, p53.

References and Notes

1. A. J. Levine, *Cell* **88**, 323 (1997).
2. Y. Haupt, R. Maya, A. Kazaz, M. Oren, *Nature* **387**, 296 (1997); M. H. G. Kubbutat, S. N. Jones, K. H. Vousden, *Nature* **387**, 299 (1997); R. Honda, H. Tanaka, H. Yasuda, *FEBS Lett.* **420**, 25 (1997).
3. J. Momand, G. P. Zambetti, D. C. Olson, D. George, A. J. Levine, *Cell* **69**, 1237 (1992).
4. C. J. Thut, J. A. Goodrich, R. Tjian, *Genes Dev.* **11**, 1974 (1997).
5. L. J. Ko, C. Prives, *Genes Dev.* **10**, 1054 (1996).
6. M. Oren, *J. Biol. Chem.* **274**, 36031 (2000).
7. C. J. Sherr, *Genes Dev.* **12**, 2984 (1998).
8. J. Roth, M. Dobbelstein, D. A. Freedman, T. Shenk, A. J. Levine, *EMBO J.* **17**, 554 (1998).
9. D. A. Freedman, A. J. Levine, *Mol. Cell. Biol.* **18**, 7288 (1998).
10. W. Tao, A. J. Levine, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6937 (1999).
11. Y. Zhang, Y. Xiong, *Mol. Cell* **3**, 579 (1999).
12. ———, *Cell Growth Differ.* **12**, 175 (2001).
13. W. Tao, A. J. Levine, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3077 (1999).
14. J. M. Stommel, *et al.*, *EMBO J.* **18**, 1660 (1999).
15. S. D. Boyd, K. Y. Tsai, T. Jacks, *Nature Cell Biol.* **2**, 563 (2000).
16. R. K. Geyer, Z. K. Yu, C. G. Maki, *Nature Cell Biol.* **2**, 569 (2000).
17. J. Lin, J. Chen, A. J. Levine, *Genes Dev.* **8**, 1235 (1994).
18. Plasmids expressing WT p53, mutant $p53^{L14Q/F19S}$, $p53^{L22Q/W23S}$, and $p53^{R273H}$ were provided by J. Chen. All other p53 mutants were generated by site-directed mutagenesis with a Quick-Change kit (Stratagene, La Jolla, CA) and verified by DNA sequencing. Cells, cell culture, and procedures for transfection, adenovirus infection, and immunoblotting are described in (*25*). Procedures for indirect immunofluorescence and heterokaryon assay are described in (*11*) except that the incubation time with primary anti-p53 was increased to overnight at 4°C to detect UV-induced p53 in MEFs. Fluorescence images were captured with a cooled charge-coupled device color digital camera (Diagnostic, model 2.2.0) and analyzed on a Macintosh computer with the public domain NIH Image program (version 1.61; available at http://rsb.info.nih.gov/nih-image/). Dilutions and sources of primary antibodies for indirect immunofluorescence are as follows: 0.2 μg/ml for mouse anti-MDM2 (clone SMP14, NeoMarkers, Fremont, CA), 0.04 μg/ml for rabbit anti-MDM2 (N-20, Santa Cruz Biotechnology, Santa Cruz, CA), 0.4 μg/ml (MEFs) or 0.2 μg/ml (other cells) for goat anti-p53 (sc-6243G, Santa Cruz), 1:5000 dilution for affinity-purified rabbit anti–Ser^{15}-phospho-p53 (#9284, New England Biolabs, Beverly, MA), 0.4 μg/ml for mouse anti-ARF (clone 14P02, NeoMarkers), and 2 μg/ml for anti-Ku (p80, clone 111, NeoMarkers). All fluorochrome-conjugated secondary antibodies (Jackson ImmunoResearch Laboratories, West Grove, PA) are diluted to 5 μg/ml. For leptomycin B treatment, cells were treated with 5 ng/ml leptomycin B for 6 hours before cell fusion and/or fixation.
19. G. S. Jimenez, *et al.*, *Nature Genet.* **26**, 37 (2000); C. Chao, *et al.*, *EMBO J.* **19**, 4967 (2000).
20. Supplementary material available on *Science* online at www.sciencemag.org/cgi/content/full/292/5523/1910/DC1.
21. W. Wen, J. L. Meinkoth, R. Y. Tsien, S. S. Taylor, *Cell* **82**, 463 (1995).
22. A. J. Giaccia, M. B. Kastan, *GenesDev.* **12**, 2973 (1998); C. Prives, *Cell* **95**, 5 (1998).
23. C. Chao, S. Saito, C. W. Anderson, E. Appella, Y. Xu, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11936 (2000).
24. S-Y. Shieh, M. Ikeda, Y. Taya, C. Prives, *Cell* **91**, 325 (1997); N. H. Chehab, A. Malikzay, M. Appel, T. D. Halazonetis, *Genes Dev.* **14**, 278 (2000).
25. Y. Zhang, Y. Xiong, W. G. Yarbrough, *Cell* **92**, 725 (1998).
26. We thank Y. Xu for providing the $p53^{L25Q/W26S}$ embryonic stem cells, S. Jones for the p53-MDM2–deficient MEF cells, C. Finlay and T. Kowalik for the MDM2 and p53 adenoviruses, and J. Chen for p53 and HDM2 plasmids. We also thank J. McCarville and G. White Wolf for technical assistance and C. Jenkins for reading the manuscript. Y.Z. is the recipient of a Career Award in Biomedical Science from the Burroughs Wellcome Fund and a Howard Temin Award from National Cancer Institute. Y.X. is the recipient of a Career Development Award from the U.S. Department of Army Breast Cancer Research Program. Supported by NIH grant CA65572 (Y.X.).

27 December 2000; accepted 19 April 2001

In Silico Mapping of Complex Disease-Related Traits in Mice

Andrew Grupe,[1]* Soren Germer,[2]* Jonathan Usuka,[3]* Dee Aud,[1] John K. Belknap,[4] Robert F. Klein,[4] Mandeep K. Ahluwalia,[2] Russell Higuchi,[2] Gary Peltz[1]†

Experimental murine genetic models of complex human disease show great potential for understanding human disease pathogenesis. To reduce the time required for analysis of such models from many months down to milliseconds, a computational method for predicting chromosomal regions regulating phenotypic traits and a murine database of single nucleotide polymorphisms were developed. After entry of phenotypic information obtained from inbred mouse strains, the phenotypic and genotypic information is analyzed in silico to predict the chromosomal regions regulating the phenotypic trait.

Identification of genetic susceptibility loci has promised insight into pathophysiologic mechanisms and the development of therapies for common human diseases. Analysis of experimental murine genetic models of human disease biology should greatly facilitate identification of genetic susceptibility loci for common human diseases. We present a computational method that markedly accelerates genetic analysis of murine disease models. A linkage prediction program scans a murine single nucleotide polymorphism (SNP) database and, only on the basis of known inbred strain phenotypes and genotypes, predicts the chromosomal regions that most likely contribute to complex traits. The computational prediction method does not require generation and analysis of experimental intercross progeny, but it correctly predicted the chromosomal regions identified by analysis of experimental intercross populations for multiple traits analyzed.

A Web-accessible database was developed, which contains allele information across 15 inbred strains and specifies genotyping assays for over 500 SNPs at defined locations in the mouse genome (http://mouseSNP.roche.com). These SNPs were identified in our laboratories by direct sequencing of polymerase chain reaction (PCR) amplification products from defined chromosomal locations. This database also incorporates published allele information for 2848 SNPs, 45% of which are characterized in a subset of *Mus musculus* strains; 55% of the SNPs are polymorphic between *Mus castaneus* and one or more *M. musculus* subspecies (*1*). User queries regarding SNPs found within a specified chromosomal region or between selected inbred strains are executed in real time and provided through a graphical user interface. The oligonucleotide primer sequences and conditions for performing allele-specific kinetic PCR genotyping assays (*2*) are also provided in the mSNP database [see supplemental material (*3*)].

To demonstrate the utility of this information, the genome of pooled DNA samples obtained from intercross progeny was analyzed by two different genotyping methods. At 16 weeks of age, the 1000 F_2 progeny of

[1]Department of Genetics and Genomics, Roche Bioscience, Palo Alto, CA 94303, USA. [2]Roche Molecular Systems, Alameda, CA 94501, USA. [3]Department of Chemistry, Stanford University, Stanford, CA 94305–5080, USA. [4]Oregon Health Sciences University and Portland Veterans Affairs Medical Center, Portland, OR 97201, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed: gary.peltz@roche.com

a C57BL/6 × B6D2 intercross display a non–sex linked, normal distribution of bone mineral density (BMD) (*4*). Phenotypically extreme F_2 progeny with the highest (n = 150 mice) and lowest (n = 149 mice) BMD (top and bottom 15%, respectively), were subjected to a whole-genome scan for association with BMD by genotyping individual DNA samples with 112 microsatellite markers. In addition, equal amounts of DNA from the high and low BMD F_2 progeny was used to form two pools of DNA samples. Allele frequencies in the pooled samples were measured for 109 SNPs found in the mSNP database with the use of the previously described allele-specific kinetic PCR method (*2*). Differences in allele frequency between the two extremes for each marker were scored. If a marker has no association with BMD, its expected frequency is 50% for both extremes. The significance of each allele-frequency difference was calculated using the z-test and plotted as a lod score (a logarithm of the odds ratio for linkage) (Fig. 1). A significant association (lod score > 3.3) was found for four regions on chromosomes 1, 2, 4, and 11 by the microsatellite and SNP genotyping methods. SNP-based genotyping identified a linkage region near the centromere of chromosome 13, which was not found using microsatellite markers. Two SNP markers (2.2 and 6.6 cM) were more proximal to the centromere of chromosome 13 than the most proximal (10 cM) microsatellite marker used for genotyping the intercross progeny. This region is being investigated with additional markers.

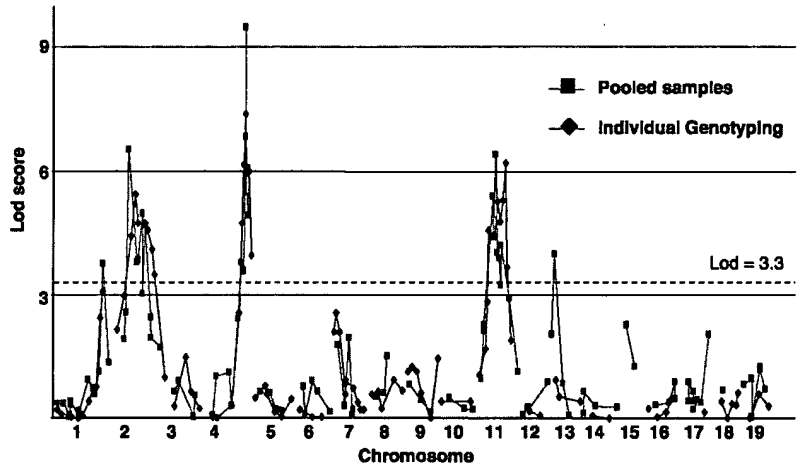


Fig. 1. Comparison of SNP-based genotyping of pooled DNA samples with microsatellite genotyping of individual DNA samples. Phenotypically extreme F_2 progeny from a B6D2 intercross with the highest and lowest BMD were subjected to whole-genome scanning for association with BMD by genotyping either individual DNA samples (from 299 mice) with 112 microsatellite markers or two pooled DNA samples (150 mice per pool) with 109 SNP markers. The significance of each allele-frequency difference was calculated using the z-test and plotted as a lod score for all chromosomes. Dashed line indicates a lod score of 3.3, the threshold for genome-wide significance.

Table 1. Comparison between experimentally identified QTL intervals with computationally predicted chromosomal regions for 10 phenotypic traits. The experimentally identified QTL intervals and phenotypic information used for computational prediction are described in the references indicated and are summarized in supplementary tables 1 and 2 (*3*). PKC, protein kinase C; Exp., total number of experimentally verified QTL intervals; Correct, number of computationally predicted regions that overlap with the experimentally verified locus; Predicted, total number of predicted regions for each phenotype; Cutoff, percentage of the mouse genome included within the computationally predicted regions.

Phenotype	Reference	Exp.	Computational		
			Correct	Predicted	Cutoff (%)
AHR	(*6, 7*)	4	4	6	15
Alcohol preference	(*12–16*)	4	3	6	16
Alcohol withdrawal	(*17*)	3	2	5	10
BMD	(*4, 18*)	7	2	2	4
Eye weight	(*19*)	1	1	4	10
Ganglion cell count	(*20*)	1	1	2	4
Lymphoma	(*21, 22*)	3	3	4	8
MHC	(*5*)	1	1	1	2
PKC activity	(*23*)	1	1	2	4
PKC content	(*23*)	1	1	6	14
Total		26	19	38	

SNP-based genotyping of pooled samples required about 20-fold fewer PCR reactions and was performed much more quickly than microsatellite genotyping of individual DNA samples. Replicate determinations (four times) were performed here to assess the reproducibility of the SNP-allele frequency determination and measurement error. On average, the standard deviation in allele frequency measurement was ±1.7%. In the future, it should be possible to reduce the number of replicate PCR assays.

We wanted to determine whether chromosomal regions regulating quantitative traits (QTL intervals) could be computationally predicted with the use of the mSNP database and available phenotypic information on inbred strains. Using the allelic distributions across inbred strains contained in the mSNP database, the computational method calculates genotypic distances between loci for a pair of mouse strains. These genotypic distances are then compared with phenotypic differences between the two mouse strains. The process is repeated for all mouse strain pairs for which phenotypic information is available. Lastly, a correlation value is derived using linear regression on the phenotypic and genotypic distances for each genomic locus.

As a first example, we used the computational method to predict the chromosomal location of the major histocompatibility complex (MHC) complex, which has been mapped to murine chromosome 17, using the known H2 haplotypes for the MHC K locus for 10 inbred strains (*5*). Phenotypic distances for strains that shared a haplotype were set to zero, and a distance of one was used for strains of different haplotypes. The SNPs within and near the MHC region had a genotypic distribution that was highly correlated with the phenotypic distances; the correlation value for this interval was 5.3 standard deviations above the average for all loci analyzed. No other peaks in the mouse genome exhibited a comparable correlation with this phenotype (Fig. 2A). This computational analysis, which required less than 1 s to run on a standard desktop computer, excluded 98% of the mouse genome from consideration without missing the genomic region known to contain the MHC.

In addition to the MHC locus, we tested the computational method using nine quantitative traits known from published studies that provided mapped QTL intervals and phenotypic data across multiple inbred strains for each trait (Table 1) (*3*). The ability of this algorithm to identify chromosomal regions regulating susceptibility to experimental allergic asthma was investigated. Analysis of

intercross progeny between susceptible (A/J) and resistant (C3H/HeJ) mouse strains identified a QTL interval on chromosome 2 and a suggested interval on chromosome 7 (*6*). Analysis of a different experimental intercross identified QTL intervals on chromosomes 10 and 11 (*7*). Phenotypic measurements for allergen-induced airway hyperresponsiveness (AHR) in four inbred strains was used for a computational genome scan. The experimentally identified QTL intervals on chromosome 2, 7, 10, and 11 were among the strongest peaks identified by the computational genome scan (Fig. 2B). The computational method excluded 85% of the mouse genome from consideration without missing the experimentally mapped QTL regions.

The ability of the computational method to correctly predict chromosomal regions containing experimentally verified QTL intervals was evaluated using 10 phenotypic traits (Table 1) (*3*). The percentage of correct predictions was characterized as a function of the percentage of the mouse genome contained within the predicted chromosomal regions. If predicted regions contained 10% of the mouse genome (by selecting 10% of the peaks with the highest correlation), then 15 of the 26 experimentally verified QTL intervals were correctly identified. As the threshold was raised, limiting the number of predicted candidate regions, more experimentally verified QTL intervals were missed. In summary, at cutoff values ranging from 2 to 16%, 19 of 26 experimentally verified QTL intervals regulating 10 phenotypic traits were correctly identified (Table 1).

We applied a Fisher Exact test to assess the significance of the computational predictions. The average size of a predicted genomic region was 38 cM, segmenting the 1500-cM mouse genome into 40 regions. Therefore, a total of 400 genomic intervals were analyzed for the 10 quantitative traits examined. At a 10% genome-wide threshold, the computational method correctly identified 15 (true positive) and missed 11 (false negative) of the 26 experimentally verified QTL intervals. The algorithm further predicted that 24 genomic intervals (false positive) contributed to a phenotypic trait where no QTL had yet been experimentally characterized, and the predictions agreed with available experimental data that 350 regions (true negative) were not QTL intervals for the 10 phenotypes examined. The Fisher Exact test yields a highly significant P value (1.0×10^{-10}), confirming significant agreement between the computationally predicted and experimentally determined chromosomal regions.

Computational analysis of the murine SNP database using phenotypic data from inbred parental strains rapidly identifies candidate QTL intervals. This can eliminate many months to years of laboratory work required to generate, characterize, and genotype intercross progeny, reducing the time required for QTL interval identification to milliseconds. In addition to its rapidity and low cost, the computational prediction method has a substantial advantage over QTL analysis using intercross progeny or recombinant inbred strains (*8*). Because it performs multiple comparisons across a range of inbred strains, the computational method takes advantage of the total genetic variation provided by available inbred mouse strains.

The ability of the computational genome scan to perform whole-genome association studies using the mouse SNP database indicates that linkage disequilibrium may extend over large regions among inbred mouse strains. Our computational results were unexpected because the number of different inbred strains for which phenotypic data was available (4 to 10) was quite limited. Positional cloning and case-control studies in human populations are routinely performed with hundreds to thousands of individuals (*9*). Several factors contribute to the successful QTL predictions by computational scanning of the mouse SNP database. The use of inbred mouse strains limits variability due to environment, and timed experimental intervention and sampling limits error in phenotypic assessment. The inbred strains are homozygous at all loci, which eliminates confounding effects due to heterozygosity found in human populations.

Recently, there has been increased emphasis on using chemical mutagenesis in the mouse as a method for studying complex

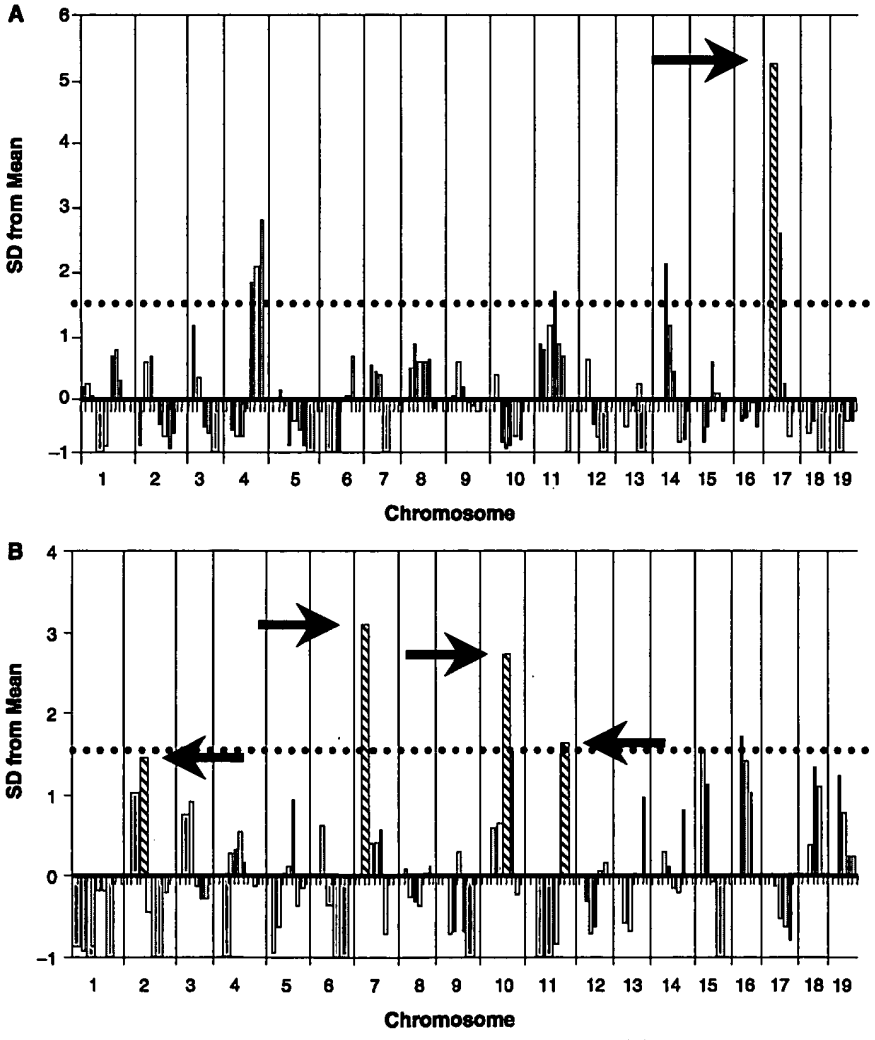


Fig. 2. Computational prediction of chromosomal regions regulating (A) MHC haplotype and (B) airway hyperresponsiveness. The correlation between the genotypic and phenotypic distributions is graphically shown for each trait; segments are arranged from centromeric to telomeric for all 19 autosomes. Each bar represents a 30-cM interval, and neighboring bars are offset by 10 cM. The dotted line represents a useful cutoff for analyzing this data; the most highly correlated 10% of the loci are above this line. Striped bars represent locations of experimentally verified QTLs.

biology. This has occurred as a result of the difficulties noted by investigators using standard methods for QTL analysis [reviewed in (*10*)]. However, these studies can be markedly accelerated by application of the genotyping method and computational tools described here. Of course, specific gene candidates must be identified to understand the genetic basis of complex disease. We have already shown how integration of gene expression data obtained with high-density oligonucleotide microarrays can be used in conjunction with the SNP genotyping method to accelerate QTL analysis (*11*). Therefore, databases with tissue-specific gene expression and phenotypic information across mouse strains could be used in conjunction with the murine SNP database to computationally identify candidate disease genes. In a hypothetical experiment, the expression of 40,000 murine genes in an affected tissue obtained from different mouse strains can be profiled. As many as 1% of the genes will be reliably demonstrated to be differentially expressed in the tissue of the mouse strains with a different phenotype. The resulting list of 400 gene candidates could be computationally reduced by 90% by searching for genes that are encoded within computationally predicted chromosomal regions, providing a reasonable starting point for analysis of complex disease biology. The application of this approach should reduce the frustrations and overcome the difficulties associated with QTL analysis in murine complex disease models.

References and Notes

1. K. Lindblad-Toh *et al.*, *Nature Genet.* **24**, 381 (2000).
2. S. Germer, *Genome Res.* **10**, 258 (2000).
3. Web figure 1, Web tables 1 and 2, and supplemental text are available at *Science* Online at www.sciencemag.org/cgi/content/full/292/5523/1915/DC1.
4. R. F. Klein *et al.*, *J. Bone Miner. Res.*, in press.
5. *JAX Notes* **475**, (1998).
6. S. Ewart *et al.*, *Am. J. Respir. Cell Mol. Biol.* **23**, 537 (2000).
7. Y. Zhang *et al.*, *Hum. Mol. Genet.* **8**, 601 (1999).
8. A. Darvasi, *Nature Genet.* **18**, 19 (1998).
9. N. Risch, K. Merikangas, *Science* **273**, 1516 (1996).
10. J. H. Nadeau, W. N. Frankel, *Nature Genet.* **25**, 381 (2000).
11. C. L. Karp *et al.*, *Nature Immunol.* **1**, 221 (2000).
12. J. K. Belknap, S. P. Richards, L. A. O'Toole, M. L. Helms, T. J. Phillips, *Behav. Genet.* **27**, 55 (1997).
13. J. A. Melo, J. Shendure, K. Pociask, L. M. Silver, *Nature Genet.* **13**, 147 (1996).
14. T. J. Phillips, J. C. Crabbe, P. Metten, J. K. Belknap, *Alcohol Clin. Exp. Res.* **18**, 931 (1994).
15. T. J. Phillips, J. K. Belknap, K. J. Buck, C. L. Cunningham, *Mamm. Genome* **9**, 936 (1998).
16. L. M. Tarantino, G. E. McClearn, L. A. Rodriguez, R. Plomin, *Alcohol Clin. Exp. Res.* **22**, 1099 (1998).
17. K. J. Buck, P. Metten, J. K. Belknap, J. C. Crabbe, *J. Neurosci.* **17**, 3946 (1997).
18. W. G. Beamer *et al.*, *Mamm. Genome* **10**, 1043 (1999).
19. G. Zhou, R. W. Williams, *Investig. Ophthalmol. Vis. Sci.* **40**, 817 (1999).
20. R. W. Williams, R. C. Strom, D. Goldowitz, *J. Neurosci.* **18**, 138 (1998).
21. M. L. Mucenski, B. A. Taylor, N. A. Jenkins, N. G. Copeland, *Mol. Cell. Biol.* **6**, 4236 (1986).
22. A. Wielowieyski, L. A. Brennan, J. Jongstra, *Mamm. Genome* **10**, 623 (1999).
23. L. D. Dwyer-Nield, B. Paigen, S. E. Porter, A. M. Malkinson, *Am. J. Physiol.* **279**, L326 (2000).

24. This work was partially supported by an NIH Genome Research Institute grant (1 R01 HG02322-01) to G.P. J.U. was partially supported by the Stanford University Genome Training Program (NIH T32HG-00044). We thank H. C. Andersen, D. Beier, D. Birch, C. Carlson, H. Erlich, U. Germer, M. Holland, and J. Sninsky for their help.

9 January 2001; accepted 7 May 2001